VERCE delivers a productive e–Science environment for seismology research

Malcolm Atkinson \textit{et al.}
School of Informatics
University of Edinburgh

IEEE 11th eScience Conference, Munich, 1 September 2015
Authors: *tip of the iceberg*

Malcolm Atkinson, Michele Carpené, Emanuele Casarotti, Steffen Claus, Rosa Filgueira, Anton Frank, Michelle Galea, Tom Garth, André Gemünd, Heiner Igel, Iraklis Klampanos, Amrey Krause, Lion Krischer, Siew Hoon Leong, Federica Magnoni, Jonas Matser, Alberto Michelini, Andreas Rietbrock, Horst Schwichtenberg, Alessandro Spinuso and Jean-Pierre Vilotte

Many more in VERCE and in other projects and organisations delivered essential parts
Outline

• What does VERCE do?
• What is its VRE for?
• How is it built?
• Lessons learned
• Summary
Supporting computational seismology
Data Intensive Research

Visualization
Data analysis / Data mining
Simulation, inversion, HR imaging

Earth's interior imaging and dynamics: noise correlation, waveform analysis
Natural hazards: new tools for monitoring earthquakes, volcanoes, and tsunami
Interaction of solid Earth with Ocean and Atmosphere: environment, climate changes

VERCE
e-Science environment for data intensive research based on an extensive service-oriented architecture

HPC/GRID Infrastructures

Cloud
Virtual Earthquake and Seismology Research Community in Europe

Virtual Environment for Earthquake Simulations and evaluation of solid-Earth Models

http://portal.verce.eu

Combined access to computing infrastructures (EGI, PRACE, Local and National Clusters), for development and execution of large HPC computations

Access and use of European data archives and services adopting International standards (FDSN, GCMT, OneGeology, EFEHR, QuakeML)

Adoption of Workflow Technologies, Data Management and Provenance System
Virtual Earthquake and Seismology Research Community in Europe

VERCE delivers a productive e-Science 1 Sept. 2015
Integrating computational and data-intensive seismology
Challenges

• Reaching *critical* thresholds
  – data: enough sources, variety, validity, accessibility and combinability
  – models: simulations and parameterisations
  – methods: established and innovative
  – organisations: cultures, facilities, ways of working
  – researchers: solid-Earth scientists collaborating intensely with computer scientists and others!
  – extensibility to cognate fields
  – flexibility to thrive as digital environment evolves
Challenges continued

• Reaching critical understandability and usability
  – making training and induction straightforward
  – productivity for “established” methods raised
  – continuous innovation of new methods

• Reaching sustainability
  – scope of science engaged
  – scope of resources contributed
  – recognition of value

• All challenges have to be achieved together
VERCE Platform: Component Interaction

Data Management

Data Archives

Science Gateway
Community Applications
VERCE Platform: Component Interaction

Data Management

Data Archives

Science Gateway Community Applications

OGC FDSN Orfeus
VERCE Platform: Component Interaction

1 - Raw data acquisition
3 - MISFIT
VERCE Platform: Component Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Data Management

Small Clusters

HPC

StgIn

Job Manager

Prov

Science Gateway Community Applications

Data Archives

OGC FDSN

<QuakeML>

Orfeus
VERCE Platform: Component Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Results and provenance management
Metadata and Provenance Archive and Services

Data Management
Web API

HPC
StgIn
StgOut
Job Manager

Small Clusters
Data Management

Data Archives

Science Gateway Community Applications

Prov

OGC® FDSN
<QuakeML>
Orfeus

iRODS
EUDAT

PRACE
egi
SCI-BUS
VERCE Platform: Component Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Data Management

HPC

StgIn

P

StgOut

Job Manager

Results and provenance management

Metadata and Provenance Archive and Services

Web API

Runtime Provenance messaging

Small Clusters

Prov

Data Archives

SCIBUS

PRACE

EGI

Science Gateway Community Applications

Interactive Validation and Visualisation throughout the process

VERCE Platform: Component Interaction
VERCE Platform: Component Interaction

1 - Raw data acquisition
2 - HPC Simulation (model stage-in)
3 - MISFIT

Data Management

Results and provenance management
Metadata and Provenance Archive and Services

Interactive Validation and Visualisation throughout the process

Science Gateway Community Applications

Small Clusters

GridFTP GridFTP GridFTP GridFTP

PROACE egi SCI-BUS

Data Archives

OGC FDSN Orfeus

VOMS X.509
Misfit Analysis
Misfit Analysis
Improving Earth models
Improving Earth models

Select region

Select stations

Choose FEM model

Choose earthquake

VERCE delivers a productive e-Science 1 Sept. 2015
Improving Earth models

Select region
- Select stations
- Choose FEM model
- Choose earthquake

Select simulation
Select HPC resources
Improving Earth models

Select region

Select stations

Choose FEM model

Choose earthquake

Select simulation

Select HPC resources

Set up and submit run
Improving Earth models

Select region

Select stations

Choose FEM model

Choose earthquake

Set up and submit run

Select simulation

Select HPC resources

Obtain traces

Prepare traces

Prepare traces

Get results

Misfit analysis

Prepare traces
Improving Earth models

Select region

Select stations

Choose FEM model

Choose earthquake

Select simulation

Select HPC resources

Obtain traces

Prepare traces

Set up and submit run

Get results

Visualise and assess

Provenance

Shared data

Misfit analysis

Prepare traces

Prepare traces

Prepare traces

Prepare traces

Prepare traces
Improving Earth models

1. Select region
2. Select stations
3. Choose FEM model
4. Choose earthquake
5. Set up and submit run
6. Get results
7. Prepare traces
8. Misfit analysis
9. Seismologist decides
10. Calculate adjoint
11. Visualise and assess
12. Provenance
13. Shared data

VERCE delivers a productive e-Science 1 Sept. 2015
Improving Earth models

Select region
- Select stations

Choose FEM model

Choose earthquake

Select simulation

Select HPC resources

Obtain traces
- Prepare traces

Set up and submit run
- Get results

Prepare traces

Visualise and assess

Misfit analysis

Seismologist decides
- Propagate adjoint
- Calculate adjoint

Provenance

Shared data
Improving method quality

Select region
- Select stations
- Choose FEM model
- Choose earthquake

Obtain traces
- Prepare traces

Set up and submit run
- Prepare traces
- Get results

Seismologist decides
- Calculate adjoint

Get results
- Visualise and assess

Provenance
- Shared data

Misfit analysis

Select simulation
- Select HPC resources

Prepare traces

Obtain traces

Prepare traces

Verce delivers a productive e-Science 1 Sept. 2015
Exploiting provenance
Architecture: flexible coupling of diverse independent systems
VERCE delivers a productive e-Science 1 Sept. 2015
Lessons learned
Respect and Harness existing cultures

• Education, working practices, technologies and organisations
  • independent and evolving
• Engage but couple flexibly
  • combine resources
  • sustain through open standards and technology
  • work to retain commitment to alignment
Agile focused task forces

• It is difficult to find multi-party ways forward
• Establish short-lived focused task forces to create a solution
  • All disciplines / viewpoints involved
  • computational seismologists, data-intensive seismologists, resource owners, data suppliers, user-interface creators, workflow experts, simulation experts, visualisation experts, system administrators
• Get help from others: EGI, EUDAT, PRACE, SCI-BUS, ...
Balance productivity with innovation

• “one or two orders of magnitude improved productivity”
  • packaged access to established methods
• Path to developing new methods
  • accessible to domain scientists without depending on ICT experts
• Path from innovation to production
  • as easy as possible without depending on ICT experts
Independent improvements

• Each organisation, sub-discipline and technology drives *their own* improvements
• Avoid composition breaking
• Pool benefits
  • continued dialogue
  • flexible software boundaries and mappings
Jam tomorrow not enough

- Vision for future facilities essential
- to head in a good direction
- But benefits today are needed
- to retain engagement
- to validate ideas
Summary and Conclusions
Simultaneous successes

• Easily used seismology VRE in production
• Computational, data and analytic power
• Balance productivity with innovation
• Integrated diversity of methods, data and resources
• Intense inter-disciplinary collaboration
• Flexible and sustainable framework
• Ready for expansion into many domains
• A platform on which EPOS will build
In Memoriam

Torild van Eck